ElSa: A Novel Real-time Wildlife Poacher Detection Solution Leveraging Machine Learning Driven Spatio-temporal Analysis of Nighttime UAV Thermal Infrared Videos

Anika Puri

Horace Greeley High School, Chappaqua, NY, USA

ElSa: A Novel Real-time Wildlife Poacher Detection Solution Leveraging Machine Learning Driven Spatio-temporal Analysis of Nighttime UAV Thermal Infrared Videos

Abstract

Wildlife poaching of endangered species such as elephants in Africa and Asia for illegal trading has become a biodiversity crisis, which has also been highlighted by the United Nations Sustainable Development Biodiversity Goal of halting biodiversity loss. Recently, unoccupied aerial vehicles (UAVs) equipped with heat-sensing infrared cameras (and coupled with computer vision software) have been deployed to help park rangers monitor protected areas at night when illegal wildlife poaching typically occurs and protected areas are closed. In order to maximize the area covered within a fixed flight time and battery constraints, the UAVs usually fly at an altitude of approximately 400 ft. Unfortunately, this results in small animal/human sizes in the captured thermal images, and consequently leads to poor detection accuracy of as low as 20% for detection of humans/potential poachers. This research leverages the spatio-temporal nature of the video data, i.e., the difference in the movement patterns of animals and humans over time, such as number of turns over time, their turning radius, speed, etc., to determine whether these features have promise in improving human vs animal classification. When tested using thermal infrared video dataset called BIRDSAI [Bondi et al., 2020], collected from four national parks in Africa, this proposed method was able to use movement pattern features for detecting humans with 81.7% accuracy. This space-time model is complemented with a new animal/human count model that leverages the herd nature of animal behavior in further improving human detection accuracy to 90.8% -a 4X improvement over current State-of-the-art results. Furthermore, a low-cost ($300) design prototype ElSa (Elephant-Savior) is demonstrated, that mitigates the need for costly high-resolution thermal cameras (costing up to $10,000), easing the burden on resource-constrained Parks in Africa.

1 Introduction

1.1 Overview

Prevention of wildlife poaching of endangered species has been highlighted by the United Nations Sustainable Development Goal SDG15 of halting biodiversity loss [UN, 2021]. The population of African elephants has fallen from an estimated 12 million a century ago to some 400,000 today. In recent years, at least 20,000 elephants have been killed in Africa each year (equivalent to one death every 26 minutes), in large part because of the illegal ivory trade - the biggest driver of elephant poaching [WWF, 2019]. African forest elephants have been the worst hit. Their population has declined by 62% between 2002-2011, with African savanna elephants declining by 30% between 2007-2014 [WWF, 2021]. This dramatic decline has continued and even accelerated with cumulative losses of up to 90% in some landscapes between 2011 and 2015. The World Wildlife Fund for Nature recently warned that unless this biodiversity crisis is addressed urgently, African elephants will become extinct within two decades [WWF, 2019]. Motivated by this crisis, wildlife conservation has become one of the most important environmental sustainability goals that also promotes economic growth in poverty-stricken regions around the world by linking development, governance and natural resource conservation to alleviate poverty [Columbia Earth Institute, 2021].

Unfortunately, elephant poaching activity continues to flourish despite a 1990 global ban [CITES, 1990] on ivory sales driven by a growing demand for products made from this material for ornaments and decorations, as well as for use in traditional Asian medicine for its purported therapeutic value. The poachers are part of powerful organized criminal networks which commonly engage in corruption, money laundering and assassinations [US Congress Report, 2012].

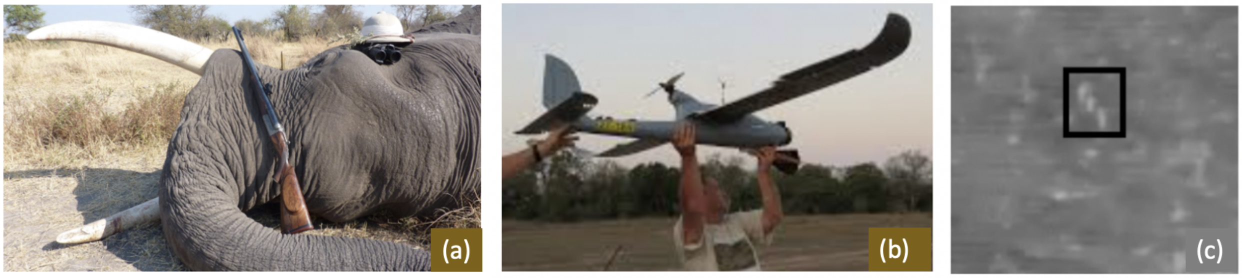


Figure 1: (a) Elephant Poaching Activity (b) Air Shepherd UAVs [Airshepherd, 2021] for wildlife conservation (c) Thermal infrared image of poachers [Bondi et al., 2020]

In order to protect wildlife from poaching [Pires and Moreto, 2016], park rangers patrol wide swaths of national parks; however, a single national park can be as large as 100,000 sq km. Part of the issue in policing these large national parks is that the governments of nations where African elephants live often lack sufficient resources to protect and monitor elephant herds, which often reside in remote and inaccessible habitats. Due to these constraints, wildlife poaching prevention efforts continue to face substantial challenges, highlighting the urgent need for a practical and cost-efficient solution to this problem.

1.2 Review of Literature

Conservation programs such as Air Shepherd [Airshepherd, 2021] have deployed unmanned aerial vehicles (UAVs) with video surveillance to protect wildlife from poaching in Africa and Asia (Figure 1(b)). Vuuren et al. presented a UAV surveillance methodology for prevention of wildlife poaching and discussed its effectiveness [Vuuren et al., 2019]. The UAV operators pre-program the drone flight path based on typical poaching hotspots and animal density which is highly correlated with poaching activity. They monitor the live video stream, transmitted via radio waves, for any signs of poachers. Should humans/potential poachers be spotted, the team manually takes control to follow the suspects, notify nearby park rangers and guide them to the poachers. Monitoring these videos all night is a difficult and painstaking task due to the quality of infrared video, an image from which is shown in Figure 1(c). With very few pixels associated with the objects of interest (humans/animals) in the UAV videos, and many objects that look similar to those of interest, hours of this live video monitoring task at night lead to human errors and missing poaching activity [Vuuren et al., 2019].

Recent progress in deep-learning and machine-learning methods has revolutionized the field of computer vision and automated object detection with applications in almost every field [LeCun et al., 2015]. Researchers have leveraged these advances in computer vision technology for wildlife conservation with automated visual surveillance for human presence with deep-learning driven object detection methods [Bondi et al., 2018b; Guo et al., 2020; Kamminga et al., 2018; López and Mulero-Pázmány, 2019; Lygouras et al., 2019]. Since most of poaching occurs at night [Wildlife Crime, 2021], high resolution thermal infrared cameras (costing over $10,000 [FLIR, 2021]) mounted on these UAVs have been deployed to detect animal and human activity in national parks in Africa and Asia [Bondi, 2018; Kamminga et al., 2018; Jiménez López and Mulero-Pázmány, 2019]. Although the resolution and cost of thermal infrared cameras have improved recently, they continue to lag their visual light cameras counterparts by an order of magnitude both in maximum resolution as well as cost, due to the complexities of the thermal sensor technologies [Embedded, 2021]. In order to maximize the surveillance area and avoid detection, the UAVs usually fly at an altitude of approximately 400 ft or above [Airshepherd, 2021]. This results in small animal/human sizes in the captured thermal videos. Due to this, it is often very difficult for even human experts to recognize poachers in these videos, leading to recognition errors. To help alleviate this problem, recent research efforts have focused on computer vision-driven automatic animal and human detection methods.

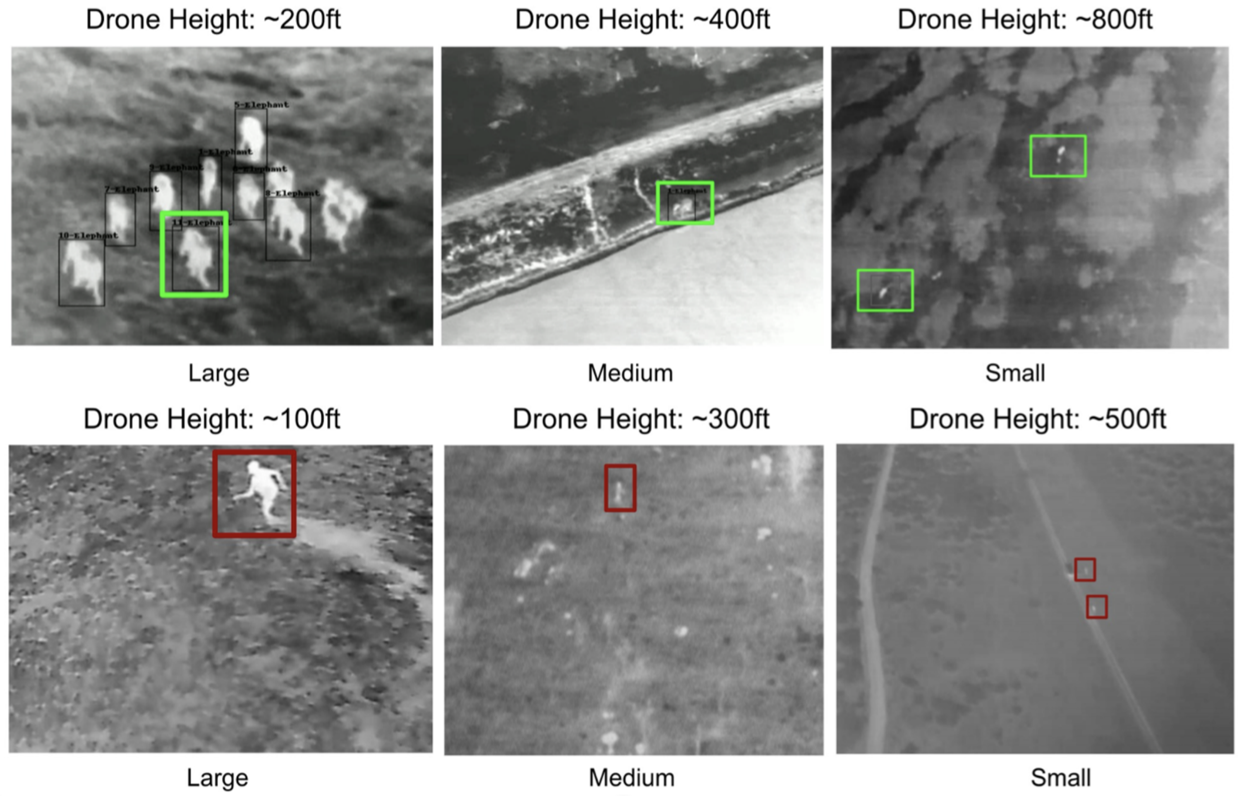
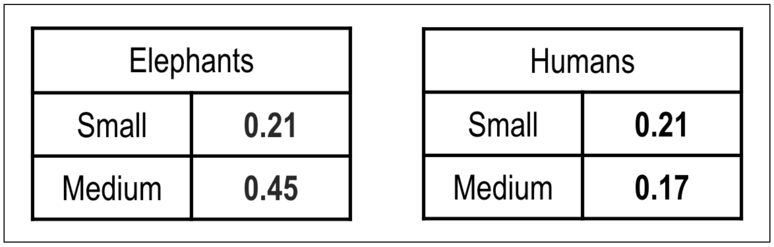


Figure 2: Decreasing size of Elephant and Poacher images with increasing UAV height from left to right, captured in thermal infrared video BIRDSAI data [Bondi et al., 2020]

In order to apply computer vision based detection, Hannaford developed EyeSpy, an application that was used by Air Shepherd in practice for detecting objects based on edge detection [Airshepherd, 2021; Hannaford, 2017]. However, several limitations prevent widespread use of this tool, such as the need for: subject matter experts for monitoring to provide parameters like edge detection thresholds, sizes, altitude, and camera look angle throughout the UAV flight. To make best use of this tool, the UAV crew either needed to restrict the way the UAV flies by keeping the flight altitude and camera look angle almost the same throughout the mission, or have the expert monitoring personnel manually adjust the parameters from time to time as the settings change. Due to this manually intensive feature engineering approach, EyeSpy has had only limited penetration in practice [Bondi et al., 2018a].



Elephants	
Small	**0.21**
Medium	**0.45**

Humans	
Small	**0.21**
Medium	**0.17**

Figure 3: Current human/animal detection (based on Object-Detection Methods) accuracy results [Bondi et al., 2020]

Bondi et al. addressed some of these limitations by leveraging a deep learning driven object detection technique using convolutional neural networks [Bondi et al., 2018a; Bondi et al., 2019]. Their method treats each frame of the video as an image, and tries to localize and recognize the objects of interest from its shape in these static images. Figure 2 illustrates the difficulty associated with distinguishing between animals vs humans objects in night time thermal infrared videos (collected by Air Shepherd UAVs in African National Parks, that are part of the BIRDSAI dataset [Bondi et al., 2020]), solely based on shape. Notice that these image frames are gray scale, with few pixels on objects of interest, i.e., humans/animals. In addition, many other objects in the video frames look similar to the objects of interest. Unfortunately, due to minute sizes of objects of interest in nighttime thermal infrared UAV videos, existing shape recognition based methods result in poor classification accuracy of as low as 20% for detecting humans, as shown in Figure 3 [Bondi et al., 2018a; Bondi et al., 2019; Bondi et al., 2020].

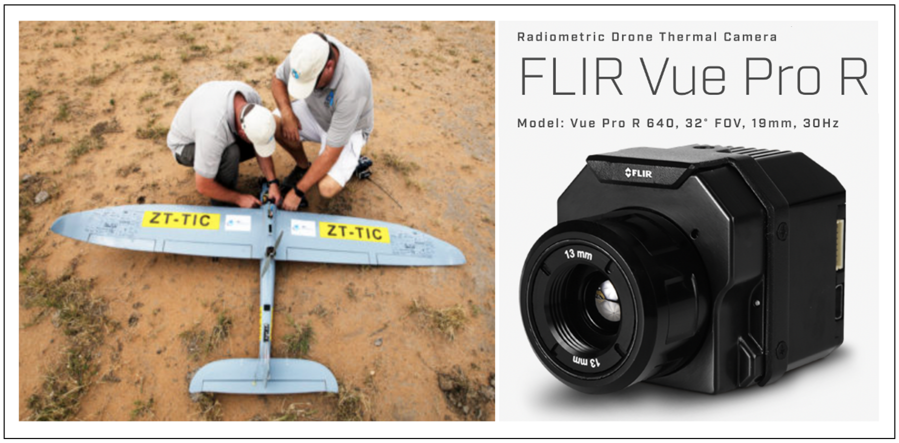

ZT-TIC
ZT-TIC
Radiometric Drone Thermal Camera
FLIR Vue Pro R
Model: Vue Pro R 640, 32° FOV, 19mm, 30Hz
FLIR
13 mm
13 mm

Figure 4: Air Shepherd UAV with Thermal Camera [Airshepherd, 2021] [FLIR, 2021].

Although a deep learning based approach is flexible but due to its poor accuracy to detect human activity in real-life nighttime thermal infrared videos, its use in practice has not progressed beyond early trials. In addition, this method is also unable to eliminate the need for high-resolution thermal cameras, costing over $10,000 each (Figure 4) [FLIR, 2021], severely burdening already resource-constrained Parks in Africa.

2 Problem Statement, Objectives, and Hypothesis

Problem Statement: *Current existing automated computer vision methods, although flexible, have poor accuracy for detecting potential poaching activity in real-life UAV nighttime thermal infrared videos. They are also unable to eliminate the need for costly high-resolution thermal cameras, a significant burden for already resource constrained parks in Africa. This poor accuracy and high-cost of thermal cameras in UAVs remains a substantial barrier to the widespread deployment of automated computer vision methods to prevent wildlife poaching.*

In order to overcome these drawbacks/problems and widely impact wildlife poaching, I formulated following **objectives** for my proposed research:

(**Obj.1**)**:** Develop a high accuracy automated computer-vision method for detection of potential poaching activity in real-life nighttime UAV thermal infrared videos.

(**Obj.2**)**:** Design and implement a low-cost prototype, built with commodity hardware, for nighttime thermal infrared video capture, and integrate with computer-vision method in *Obj*.1, for high accuracy detection of human/potential poacher activity.

Fortunately, video data carries much more information than static image frames. It has time domain information which captures the spatio-temporal movements of objects of interest. Unfortunately, current human detection methods based only on object detection in images [Bondi et al., 2018a; Bondi et al., 2020] fail to leverage this important spatio-temporal nature of the video data. Studies show that the movement patterns of elephants differ significantly from those of humans with respect to speed, turning patterns, etc. [Wilson et al., 2015; de Knegt et al., 2021; Ren and Hutchinson, 2008]. In addition, it is well known that several animal species such as elephants exhibit herd behavior, i.e., they are typically found in groups [Sumpter, 2010]. Based on these insights, I developed following **hypotheses** for my proposed research :

(**H1**)**:** Differences in animal and human spatio-temporal movement patterns in real-life UAV thermal infrared video data can be leveraged to substantially increase the accuracy of identifying human/poacher activity in wildlife national parks.

(H2): Animal herd behavior can be leveraged to further improve human and animal detection accuracy in real-life infrared video data.

In the following, the overall methodology for achieving my research objectives is discussed in detail. This methodology builds upon the insights captured as part of the hypotheses above.

3 Methodology

Over the course of my research, I have been responsible for the discovery and conceptualization of the problem, background info/research, outlining goals and hypotheses, developing and executing methodology, coding and analyzing the machine learning model, analyzing results, formalizing conclusion, and researching next steps. I am thankful to my mentor for encouragement and feedback throughout my journey as a researcher.

3.1 Human vs Animal Machine Learning Model training methodology

Section 3.1 details the overall methodology to achieve the research objective *Obj*.1. The implementation methodology for achieving research objective *Obj*.2 is elaborated in Section 3.2.

3.1.1 Extracting spatio-temporal series training data

The first step in testing my hypothesis $H1$ was to curate spatio-temporal movement patterns of humans and elephant present in a real-life UAV thermal infrared video dataset (ground truth data), The Benchmarking Infrared Dataset for Surveillance with Aerial Intelligence (BIRDSAI) was used, which is a long-wave thermal infrared (TIR) dataset [Bondi et al., 2020] containing real-life nighttime videos of animals and humans in Southern Africa, taken from UAVs flying at various heights. This dataset is used for human/animal detection machine learning human/animal detection model training. It includes TIR videos of humans and animals with several challenging scenarios like scale variations, background clutter due to thermal reflections, large camera rotations, and motion blur.

Fortunately, BIRDSAI dataset had a large number of video frames with elephant sightings, which yielded good amount of data to train a classification model for identification of elephant movements among animals as an exemplar, along with humans/potential poachers.

In order to extract spatio-temporal patterns, movement with respect to the fixed object and human/elephant ground truth label positions in the BIRDSAI dataset [Bondi et al., 2020] is mea-

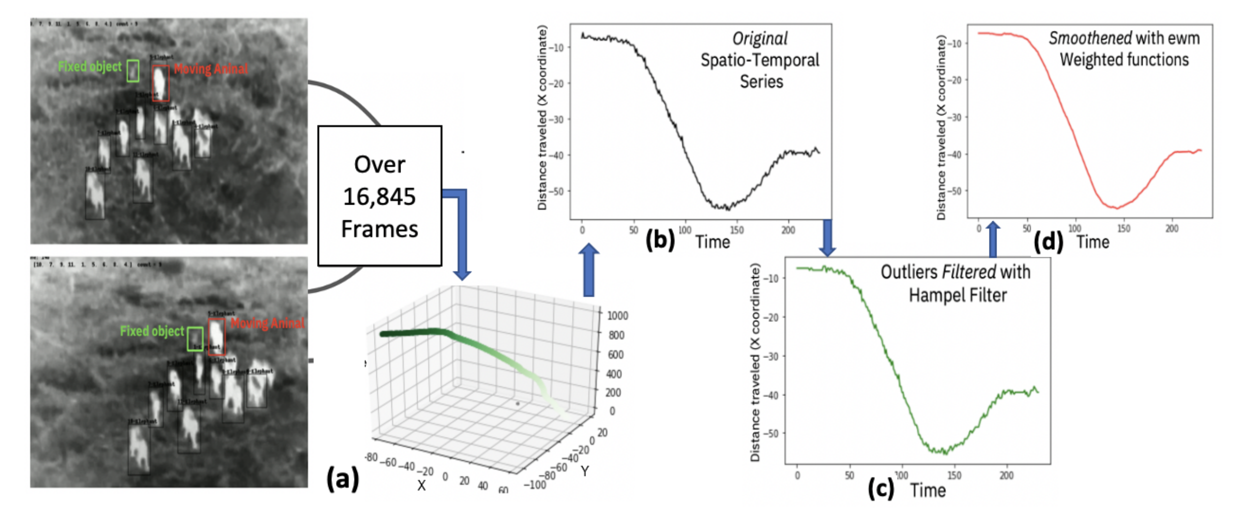


Figure 5: (a) Extracting spatio-temporal movement series from TIR videos (b) Original spatio-temporal movement (b) Outliers filtered w/ Hampel Filter (c) Smoothening w/ *ewm* weighted functions.

sured. The fixed object selected manually was either a tree or a water's edge (prominent in TIR videos). The Discriminative Correlation Filter with Channel and Spatial Reliability (DCF-CSR) also known as CSRT tracker, was used to track the fixed object due to its higher precision and performance in tracking objects of interest in thermal infrared videos [AlMansoori et al., 2020]. The results of tracking fixed objects through the CSRT tracker was also verified through extensive manual inspection as well. For every instance of a human/elephant, several fixed objects were tracked to maintain the relative movement data of the object of interest in case one of the fixed objects went out of video frame. The objects of interest were initialized in the first frame using the ground truth labels in the BIRDSAI dataset and were then dynamically tracked with a CSRT tracker along with the fixed object tracking (also with CSRT tracker). This data was used to derive the spatio-temporal time series of animal/elephant movement by computing the relative distance between the tracking coordinates of the fixed object and the object of interest. This process is repeated for both humans and elephants, and yields high quality raw data of spatial movement over time, captured as a time series, as shown in Figure 5(a)-(b).

3.1.1.1 Outlier Removal and time-series smoothing Some jitter due to the imperfection of the object tracking algorithm can naturally creep into the movement patterns. Since it is physically impossible for either humans or elephants to move tens of meters within a fraction of a second (with respect to the fixed object), these outliers were removed in the spatio-temporal time series with a Hampel filter [Liu et al., 2004], which detects points with significant deviation from the sliding window median and replaces the detected outliers with the median. Figure 5(a)-(b) shows a representative spatio-temporal series, and the filtered series with outliers removed is shown in Figure 5(c). This time series can be further smoothed to remove the unnatural minor jittery patterns seen in Figure 5(c): the exponential weighted functions from pandas *ewm* library in python was used [McKinney et al., 2011] to smooth out these minor jitter movement patterns while maintaining their overall movement features. The smoothed representative spatio-temporal movement pattern is shown in Figure 5(d).

These curated spatio-temporal series derived for numerous instances of both humans and elephants serve as the training data for the automated human/elephant classification model which was used to test hypothesis $H1$.

3.1.2 Training KNN classification model by extracting features from the spatio-temporal series

In the proposed methodology, k-nearest neighbor (KNN) machine learning algorithm is used for training a classification model. KNN is simple, easy to interpret, and works well on small amounts of training data. Typical training data for a KNN classification model includes several features that characterize the unique nature of the spatio-temporal series derived above. For the training data derived above, these features are, speed and duration, number of turns, and their radius, for each of this time series. These features are automatically extracted from each of the time series using the Ramer–Douglas–Peucker algorithm [Ramer, 1972] as follows.

The Ramer–Douglas–Peucker (RDP) Algorithm [Ramer, 1972] is most commonly used in geospatial visualizations. The main purpose of this algorithm is to find a similar curve with fewer points for a given curve composed of line segments (called Polylines). In order to use this algorithms, the user specifies a threshold limit ϵ, that is used to determine which points can be discarded for approximation. This algorithms stores the start and the ending point of the movement path and then it draws the shortest line from these bookend points. It then determines the point farthest away from the line segment with the first and last points as new end points. Any points within the ϵ distance (a pre-specified parameter) from this line will be removed and the approximation will be redrawn. This process repeats recursively until the new approximation for the polyline has been formed. The simplified curve consist of a subset of points that defined the original curve. A representative movement path for a poacher is shown in Figure 6(a) and the turning points on that

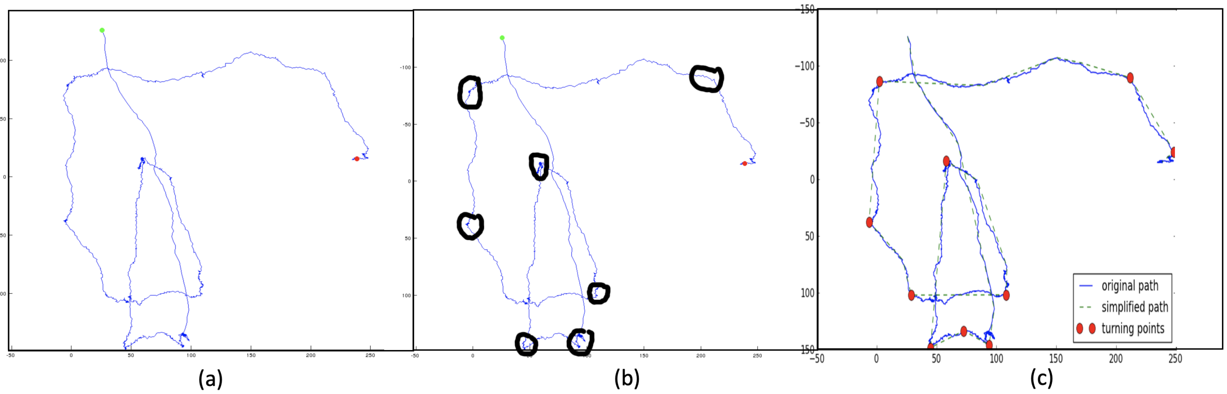
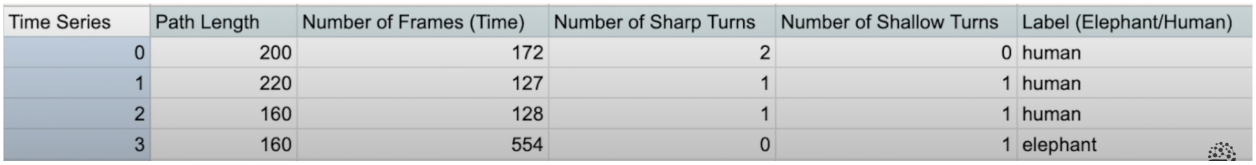


Figure 6: Extracting number of turning points from spatio-temporal movement series (a) Original spatio-temporal movement path after smoothing (b) Manually annotated turning points (c) Turning points (red dots) computed with Ramer–Douglas–Peucker algorithm along with the approximate path/polylines (dotted).

path to determine number of turns are manually annotated in Figure 6(b). The results of running the RDP algorithm for automatically extracting the number of and location of turning points on a representative human movement path is given in Figure 6(c).

RDP algorithm enables the extraction of number of turns, and the nature of those turns (shallow vs sharp turn), for each of the spatio-temporal time series derived in previous Section 3.1.1.1. Figure 7 shows representative extracted feature data for few human and elephant spatio-temporal series.



Time Series	Path Length	Number of Frames (Time)	Number of Sharp Turns	Number of Shallow Turns	Label (Elephant/Human)
0	200	172	2	0	human
1	220	127	1	1	human
2	160	128	1	1	human
3	160	554	0	1	elephant

Figure 7: Representative extracted feature data for few human and elephant spatio-temporal series for training of the KNN classification model.

These extracted features along with their corresponding labels (human/elephant) yield the ground truth data for training the human vs elephant K-Nearest Neighbor automated classification model.

3.1.3 Training KNN classification model by using dynamic time warping similarity metrics

The curated spatio-temporal series derived in Section 3.1.1.1, have some unique characteristics. They are multivariate in nature due to movement over time in both horizontal and vertical dimensions. In addition, they are also of unequal length as they are extracted from real infrared videos (ground truth data from BIRDSAI dataset [Bondi et al., 2020]). In the past, the only robust way to build a classification model for such series was to extract the features as discussed in Section 3.1.2. However, recent progress in machine learning methods for time-series has enabled classification of unequal, multivariate time series [Löning et al., 2019; Tavenard et al., 2020].

Methods such as dynamic time warping (DTW) [Berndt and Clifford, 1994] are being deployed to find patterns of interest in complex time series data. New techniques combining DTW driven distance metrics with various clustering methods are enabling wider adoption of machine learning for real-world time series data [Pouw and Dixon, 2020; Ten Holt et al., 2007]. A newly observed time series is matched to the nearest known time series instance and the known instance's class is checked - if it is a human, then the newly observed instance is also likely to be a human. The "distance" between two time-series can be measured with a euclidean distance metric, or a dynamic time warping metric. For euclidean distance metric, series have to be exactly the same for them to have a relationship. In contrast, Dynamic Time Warping metric can take speed, time, and patterns of the series of unequal lengths into account (Figure 8).

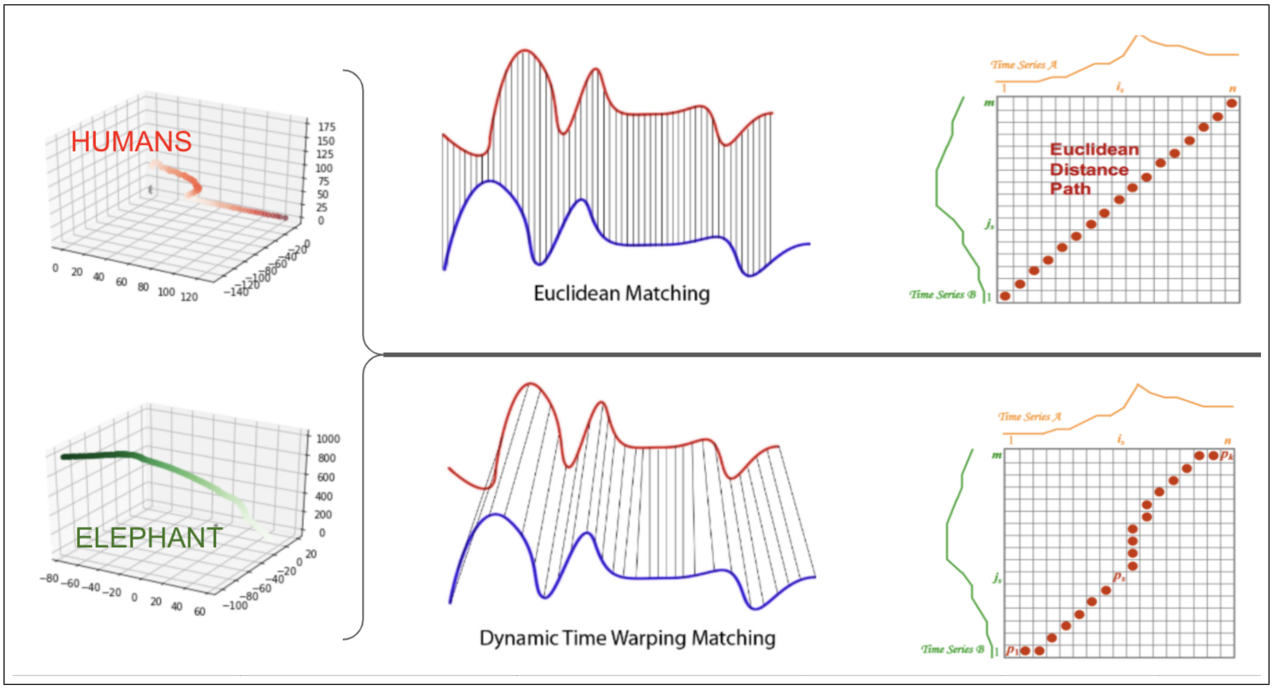


Figure 8: Comparing human/elephant spatio-temporal series with different distance metric for KNN algorithm.

For training the human/elephant detection model [Puri, 2021], KNN clustering algorithm implementation *KNeighbors* in the *TimeSeriesClassifiers* function of the recently released *tslearn* library [Tavenard et al., 2020] was used along with dynamic time warping driven similarity/distance metric.

3.1.4 Herd Count Model

While analyzing and curating the human/elephant movement time-series training data, it was empirically observed that animals such as elephants were usually present in groups/herds in BIRDSAI videos. This led to the hypothesis $H2$ that the number of objects of interest in a frame could be used as an important feature for further improving the accuracy of human/elephant prediction. The features extracted in Section 3.1.2 were further enhanced with an additional feature, "object count," that captures this herd behavior in the corresponding video segment. The spatio-temporal KNN model was retrained with this enhanced feature set with the goal of enhancing classification accuracy of detecting humans in nighttime infrared thermal videos. A *herd count model* was next developed from the BIRDSAI ground truth data.

3.2 Real-time Inference Methodology for detection of potential poachers and Implementation of Hardware and Software Design Prototype

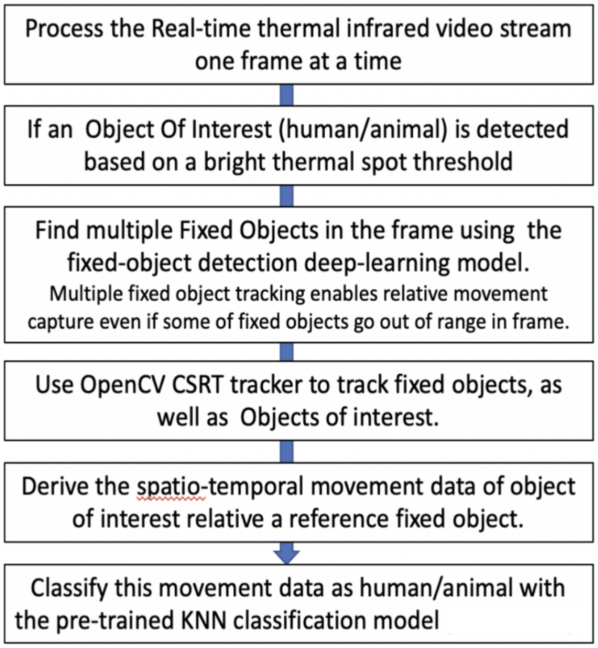

Process the Real-time thermal infrared video stream one frame at a time
If an Object Of Interest (human/animal) is detected based on a bright thermal spot threshold
Find multiple Fixed Objects in the frame using the fixed-object detection deep-learning model.
Multiple fixed object tracking enables relative movement capture even if some of fixed objects go out of range in frame.
Use OpenCV CSRT tracker to track fixed objects, as well as Objects of interest.
Derive the spatio-temporal movement data of object of interest relative a reference fixed object.
Classify this movement data as human/animal with the pre-trained KNN classification model

Figure 9: Real-time detection/inference methodology.

In order to achieve objective $Obj.2$, a real-time human/poacher detection method is needed to detect human activity in real-time thermal infrared (TIR) video feed during UAV's flight. Such a workflow is given in Figure 9. To classify a movement pattern as human/elephant in real-time, identification and tracking of fixed objects along with objects of interest is required. For this purpose, a dataset was curated with over 1,000 TIR images of trees, bushes, and water's edge as training data, which are almost universally present in almost every video frame. A fixed object detection model was then trained with Google AutoML Vision [Bisong, 2019] capability which uses transfer learning and neural architecture search to tune the final layers of neural network and customizes the trained model for the given labeled data. It took approximately 2 hrs of training time with AutoML's default compute resources [Bisong, 2019] (2 Google Tensor Processing Units). The AutoML object recognition computer vision model thus derived can detect several fixed objects with high precision (83.3%), based on the prior manually-verified tracking labels. This AutoML model enables extraction of object of interest (human/elephant) spatio-temporal movement pattern in real time.

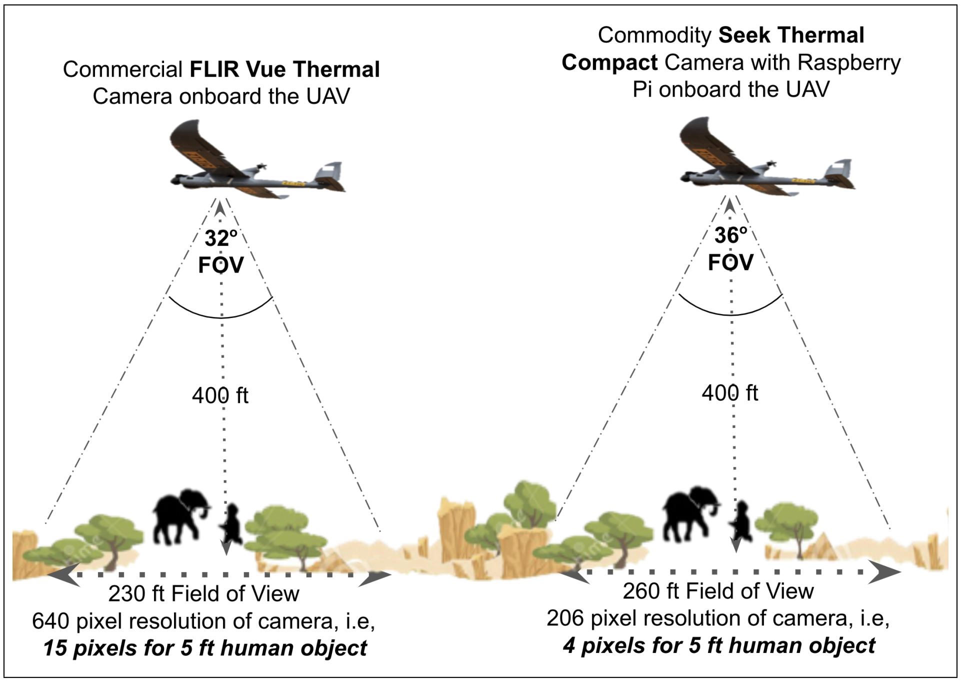

Commercial **FLIR Vue Thermal** Camera onboard the UAV
32° FOV
400 ft
230 ft Field of View
640 pixel resolution of camera, i.e,
15 pixels for 5 ft human object
Commodity **Seek Thermal Compact** Camera with Raspberry Pi onboard the UAV
36° FOV
400 ft
260 ft Field of View
206 pixel resolution of camera, i.e,
4 pixels for 5 ft human object

Figure 10: Comparison of commercial vs commodity Thermal camera onboard the UAV, and field of vision at 400ft and its ground coverage.

As discussed earlier, currently, Air Shepherd UAVs in operations [Airshepherd, 2021] require an expensive thermal infrared FLIR camera [FLIR, 2021] with a resolution of 640 x 512 pixels and a field of vision of 32^o costing over $10,000. This yields very high resolution of over 327,000 pixels, which is required for any shape based object detection method. Basic geometry calculations convey

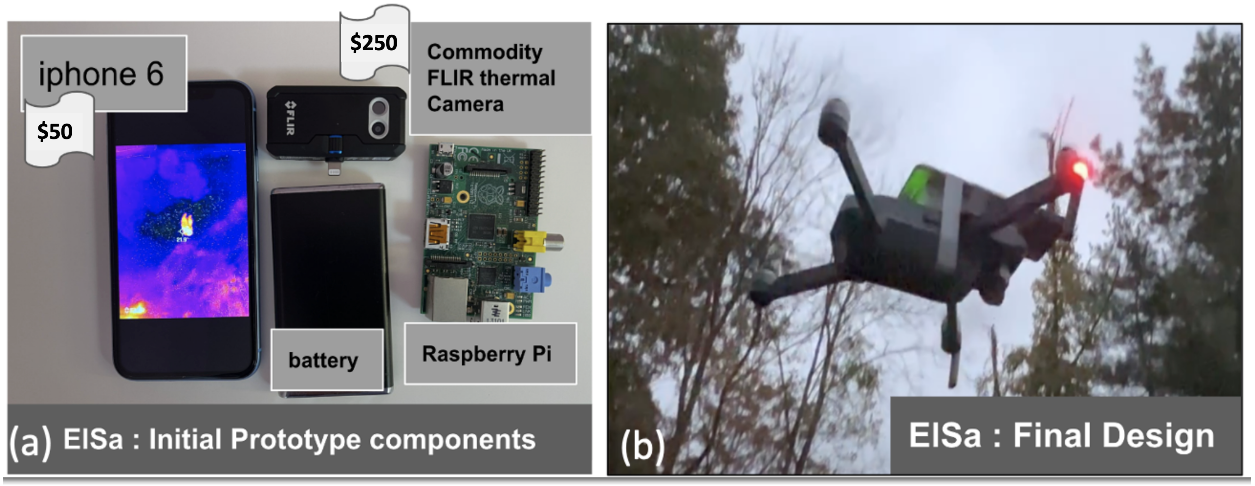

iphone 6
$50
$250
Commodity FLIR thermal Camera
battery
Raspberry Pi
(a) ElSa : Initial Prototype components
(b)
ElSa : Final Design

Figure 11: (a) Hardware Design prototype components (b) Design Prototype with a UAV

that a 5ft human object of interest will map to 15 pixels in the video frame of the commercial FLIR thermal camera as shown in Figure 10. In contrast, a commodity Camera ”FLIR pro one” [FLIR, 2021] available for less than $250 will yield 31,000 pixels for the video frame. With this commodity camera, the 5ft human object of interest will map to 4 pixels in its video frame as shown in Figure10. Since our proposed human/potential poacher activity detection method relies on spatio-temporal movement patterns as opposed to shape based object detection, it mitigates this dependence on high-resolution/image quality of the object. This enables the design and development of a low-cost prototype *ElSa* (*El*ephant *Sa*vior) which can be built with commodity hardware along with integration of our real-time inference methodology discussion in Figure 9.

Figure 11(a) shows ElSa’s design prototype which is assembled from such commodity hardware: A commodity ”FLIR Pro one thermal camera with 206x156 pixel resolution ($250) and an off the shelf $50 iphone6. For real-time processing, the inference methodology and algorithm in Figure 9 were implemented in an application in swift programming language on the iphone6 with Xcode development environment [Apple-Xcode, 2021]. The software code for the ElSa’s poacher detection methodology is open-sourced in a github repository [ElSa github, 2021].

The following section discusses the results of this proposed methodology on a real-life BIRDSAI dataset, enabling robust testing of research hypotheses $H1$-$H2$.

4 Results and Discussion

4.1 Results of Curating real-life BIRDSAI dataset into spatio-temporal series

The Benchmarking Infrared Dataset for Surveillance with Aerial Intelligence (BIRDSAI) [Bondi et al., 2020] contains over 162,000 frames of nighttime thermal videos labeled small, medium, large representing the surveillance footage with UAVs at various heights in Southern Africa National Parks. In these videos, any human activity during the night is presumed by park rangers to be suspicious poaching activity warranting further investigation. Since these videos captured real-life surveillance footage, there were lot more frames with elephants (16,845 frames) than with humans/poachers (1,853 frames). Curation of this Real-life thermal infrared video data using the proposed methodology discussed in Section 3.1.1), including rotating the spatio temporal series in 4 random directions (data augmentation for model robustness) resulted in 516 spatio-temporal movement series, out of which 408 series were for elephants - the animal in majority of images, and 108 series were for humans (Figure 12). These series were then processed for outlier removal and smoothened (as discussed in Section 3.1.1.1). For Hampel filter algorithm, a threshold of 3 standard deviations and a windows size of 5 was used to filter these abrupt unnatural movements.

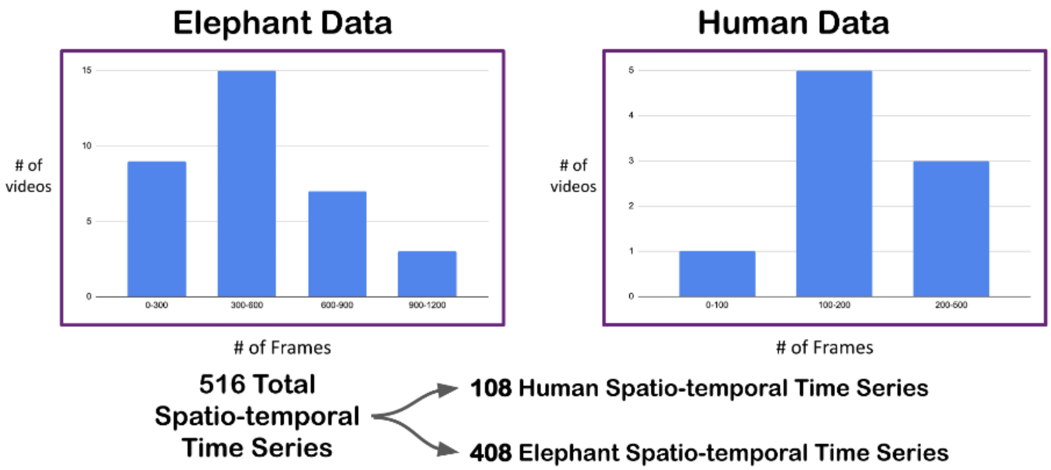


Figure 12: Curated BIRDSAI spatio-temporal series data distribution.

Among the curated spatio-temporal data derived above, 372 times series were used for training the KNN model as discussed in Section 3.1 (300 time series for elephant movements, and 72 time series for human movement). In order to make the model more robust, The remaining 144 time series (i.e., approx. 30% of the overall data) were excluded from training set, i.e, 108 for elephant movements and 36 for human movements were reserved for testing.

4.2 Herd Count Model Results

The results of the herd count model discussed in Section 3.1.4 are given in Figure 13, which shows that as the number of objects of interest in a frame increases the probability of the object of interest being a human decreases, while it being a elephant increases. As shown in Figure 13, the cross over occurred for the number objects of interest being greater than 4 in the frame which becomes the herd_threshold count parameter during the real-time inference flow given in Figure 5.

This supports the hypothesis $H2$ that herd object_count can be added as an important feature to the training feature set to enhance the accuracy of KNN classification model for distinguishing between humans vs elephants.

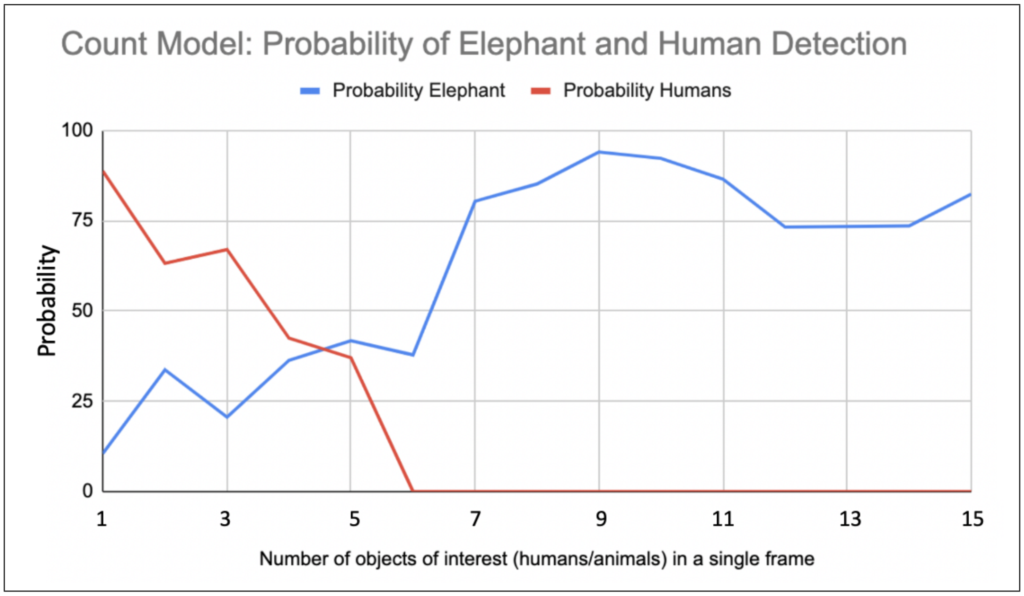


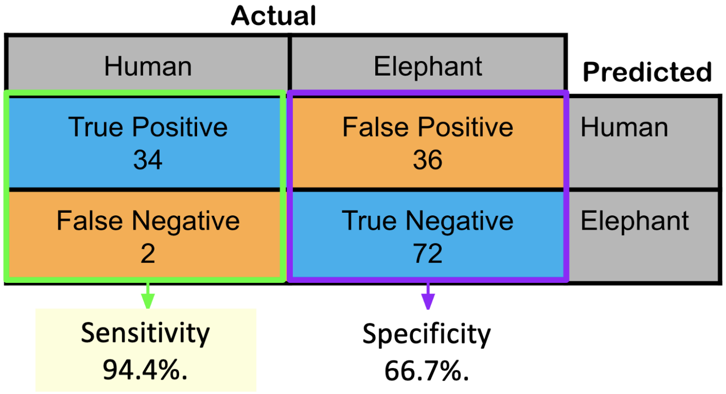
Figure 13: Herd Count Model Results.

4.3 Results of KNN classification model training with extracted features

Following the curation of 516 spatio-temporal movement series, features (speed and duration, number of turns, and the nature of those turns -shallow vs sharp turn (as shown with few representative samples in Figure 7) were extracted from each of the series with RDP algorithm as discussion in Section 3.1.2. Among these, the feature vectors of training dataset (derived in Section 4.1 earlier) were used to train the human/elephant classification KNN model (discussed in detail in Section 3.1.2. Since detecting human activity is of highest importance for prevention of wildlife poaching, sensitivity of human detection was prioritized in the model training.

When the trained model was used to classify the movement as human or elephant in this test set, *spatio-temporal KNN model achieved an accuracy of 81.7%.* Adding the herd model features to the spatio-temporal model training increases the accuracy of human/elephant prediction to 90.8%. Furthermore, achieving a precision score of 1.0 for human prediction: detecting any suspicious human activity with high precision is critical in a wildlife conservation scenario. For the human activity prediction, recall was 0.5, yielding an F1-score of 0.67. For prediction of the elephant movements, the model achieved a precision score of 0.9 and a recall of 1.0 - corresponding to an F1-score of 0.95.

4.4 Results of KNN classification model training by leveraging Dynamic Time Warping as Similarity Metrics



Actual: Human	Actual: Elephant	Predicted
True Positive 34	False Positive 36	Human
False Negative 2	True Negative 72	Elephant
Sensitivity 94.4%.	Specificity 66.7%.	

Figure 14: KNN Spatio-temporal classification Model Training Results by leveraging Dynamic Time Warping similarity metrics during KNN training.

After training the proposed model with the methodology in 3.1, the dynamic time warping KNN model was tested with these 144 spatio-temporal series (testset) it had never seen before. The DTW-KNN model was able to automatically classify 34 out of 36 human spatio-temporal series correctly, i.e., a 94.4% sensitivity for human activity detection. Although prioritized lower, the spatio-temporal model also correctly classified 72 out of 108 elephant movement series as well, yielding an specificity of 66.7%. These results are shown in Figure 14.

These results on the real-life BIRDSAI dataset support the hypothesis $H1$ that this proposed research can very efficiently utilize the difference in elephant and human movement patterns in UAV thermal infrared video data to significantly increase the accuracy of identifying human/poacher activity (achieving over 90% sensitivity -a 4X accuracy improvement over existing methods) in wildlife national parks, fulfilling the research objective *Obj.*1 as well.

4.5 ElSa's Design Prototype with field testing

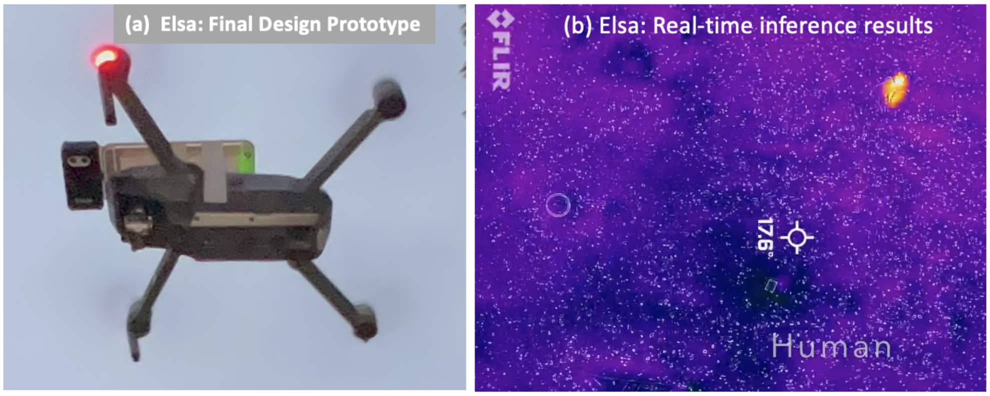

(a) Elsa: Final Design Prototype
(b) Elsa: Real-time inference results
FLIR
17.6°
Human

Figure 15: ElSa's Hardware and Software Design prototype: (b) Final design in action (b) Real-time inferencing results

Section 3.2 discussed the design and development of ElSa's prototype and its integration with the spatio-temporal human/elephant classification model deployed on an iphon6 as an app with a commodity thermal camera for real-time human activity recognition. We coupled ElSa's prototype to a DJI mavic Pro [DJI Mavic Pro, 2021] commodity UAV/drone as shown in Figure 15(a). With a commercial drone pilot license I was able to test this combined prototype for feasibility in my yard, flying it 400 ft. above the ground. A static output of real-time inference results of ElSa's integrated hardware and software design prototype are shown in Figure 15(b).

These design prototype results along with the accuracy of the human/elephant detection model discussed above further demonstrate that this solution enables significant cost saving: enabling the use of a much cheaper onboard UAV thermal camera ($10,000 vs $300) for real-time human/potential poacher detection in real-life thermal infrared videos, achieving objective *Obj.*2 of the proposed research.

5 Future Research

The BIRDSAI dataset had a large number of UAV thermal infrared video frames with elephant sightings. This yielded a good amount of data to train the proposed high-accuracy machine learning classification model for identification of elephant movements, along with humans/potential poachers. However, numerous other animal species face similar biodiversity crisis levels as elephants. In the 1970s, there were about 70,000 black rhinos in Africa, and today fewer than 5,000 are left in the wild. These animals play a crucial role in preserving our planet's biodiversity, helping to maintain healthy habitats for many other species. Enhancing the training data to capture a diversity of animal types is an important part of future research. This enhanced training dataset can further improve the impact of our human/animal classification model and ensuring that ElSa can be used to protect a broad array of endangered species. In addition, I will continue to test and tune ElSa's hardware and software to make it more robust and accurate with testing in a diversity of terrains and environments. I have already obtained a United States FAA commercial drone pilot license in August'2021 to focus on this goal. Following my paper presentation [Puri and Bondi, 2021] at the ACM Knowledge Discovery and Data Mining (KDD) Conference, Fragile Earth Workshop, I am in touch with AI for Good organization (sponsors of the workshop) to further scale my research.

6 Conclusion

Poaching of elephants and other endangered species in Africa has reached crisis proportions, which is also highlighted by UN Sustainability goal of halting biodiversity loss. Recently, UAVs equipped with heat-sensing infrared cameras coupled with computer vision software have been deployed to help park rangers monitor protected areas at night when illegal wildlife poaching typically occurs. Unfortunately, current existing automated computer vision methods, although flexible, have poor accuracy for detecting potential poaching activity in real-life UAV nighttime thermal infrared videos. They are also unable to eliminate the need for costly high-resolution thermal cameras, a substantial burden for already resource constrained parks in Africa. In this research, a novel spatio-temporal model that significantly improves Human vs Animal Detection accuracy in thermal infrared drone videos for prevention of wildlife poaching is proposed. When tested on a real life night time infrared videos dataset, collected from four national parks in Africa, this method was able to detect poachers with over 90% accuracy - a 4X improvement over the existing state-of-the-art methods. To the best of our knowledge, this is the first method [Puri and Bondi, 2021; Puri, 2021] that utilizes the animal and human movement patterns for significantly improving poacher detection accuracy in infrared thermal wildlife video data. Since this solution eliminates the need for shape based object detection algorithms used by existing methods, it enables the use of commodity thermal cameras costing <$250, as opposed to commercial high-resolution nighttime-thermal cameras costing over $10,000, as shown in ElSa's design prototype along with its real-time inference results. This novel high accuracy real-time wildlife poacher detection solution leveraging machine learning driven Spatio-temporal analysis has the potential to save thousands of endangered animals, a significant contribution to the UN Sustainability Development Biodiversity goal.

References

[Airshepherd, 2021] Airshepherd (2021). Airshepherd: The lindbergh foundation. http://airshepherd.org.

[AlMansoori et al., 2020] AlMansoori, A. A., Swamidoss, I., Sayadi, S., and Almarzooqi, A. (2020). Analysis of different tracking algorithms applied on thermal infrared imagery for maritime surveillance systems. In *Artificial Intelligence and Machine Learning in Defense Applications II*, volume 11543, page 1154308. International Society for Optics and Photonics.

[Apple-Xcode, 2021] Apple-Xcode (2021). Apple xcode integrated development environment, https://developer.apple.com/documentation/xcode.

[Berndt and Clifford, 1994] Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA:.

[Bisong, 2019] Bisong, E. (2019). Google automl: cloud vision. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, pages 581–598. Springer.

[Bondi, 2018] Bondi, E. (2018). Ai for conservation: Aerial monitoring to learn and plan against illegal actors. In *IJCAI*, pages 5763–5764.

[Bondi et al., 2018a] Bondi, E., Fang, F., Hamilton, M., Kar, D., Dmello, D., Choi, J., Hannaford, R., Iyer, A., Joppa, L., Tambe, M., et al. (2018a). Spot poachers in action: Augmenting conservation drones with automatic detection in near real time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

[Bondi et al., 2019] Bondi, E., Fang, F., Hamilton, M., Kar, D., Dmello, D., Noronha, V., Choi, J., Hannaford, R., Iyer, A., Joppa, L., et al. (2019). Automatic detection of poachers and wildlife with uavs. *Artificial Intelligence and Conservation*, 77.

[Bondi et al., 2020] Bondi, E., Jain, R., Aggrawal, P., Anand, S., Hannaford, R., Kapoor, A., Piavis, J., Shah, S., Joppa, L., Dilkina, B., et al. (2020). Birdsai: A dataset for detection and tracking in aerial thermal infrared videos. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1747–1756.

[Bondi et al., 2018b] Bondi, E., Kapoor, A., Dey, D., Piavis, J., Shah, S., Hannaford, R., Iyer, A., Joppa, L., and Tambe, M. (2018b). Near real-time detection of poachers from drones in airsim. In *IJCAI*, pages 5814–5816.

[CITES, 1990] CITES (1990). Convention on international trade in endangered species ivory ban, https://www.hsi.org/news-media/african_ivory_trade/.

[Columbia Earth Institute, 2021] Columbia Earth Institute (2021). Columbia university climate school - the earth insititute, https://www.earth.columbia.edu/articles/view/2580.

[de Knegt et al., 2021] de Knegt, H. J., Eikelboom, J. A., van Langevelde, F., Spruyt, W. F., and Prins, H. H. (2021). Timely poacher detection and localization using sentinel animal movement. *Scientific reports*, 11(1):1–11.

[DJI Mavic Pro, 2021] DJI Mavic Pro (2021). https://www.dji.com/mavic.

[ElSa github, 2021] ElSa github (2021). Anika puri elsa github software repository, https://github.com/anikapuri/tech-for-wildlife-conservation.

[Embedded, 2021] Embedded (2021). Reducing size, power, and cost for infrared thermal imaging applications, Embedded Magazine.

[FLIR, 2021] FLIR (2021). FLIR 640 vue pro thermal infrared imaging camera, https://www.flir.com/products/vue-pro-r/?model=436-0024-00s.

[Guo et al., 2020] Guo, R., Xu, L., Cronin, D., Okeke, F., Plumptre, A., and Tambe, M. (2020). Enhancing poaching predictions for under-resourced wildlife conservation parks using remote sensing imagery. *arXiv preprint arXiv:2011.10666*.

[Hannaford, 2017] Hannaford, R. (2017). personal communication.

[Jiménez López and Mulero-Pázmány, 2019] Jiménez López, J. and Mulero-Pázmány, M. (2019). Drones for conservation in protected areas: present and future. *Drones*, 3(1):10.

[Kamminga et al., 2018] Kamminga, J., Ayele, E., Meratnia, N., and Havinga, P. (2018). Poaching detection technologies—a survey. *Sensors*, 18(5):1474.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

[Liu et al., 2004] Liu, H., Shah, S., and Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & chemical engineering*, 28(9):1635–1647.

[Löning et al., 2019] Löning, M., Bagnall, A., Ganesh, S., Kazakov, V., Lines, J., and Király, F. J. (2019). sktime: A unified interface for machine learning with time series. *arXiv preprint arXiv:1909.07872*.

[López and Mulero-Pázmány, 2019] López, J. J. and Mulero-Pázmány, M. (2019). Drones for conservation in protected areas: Present and future. drones 3, 1.

[Lygouras et al., 2019] Lygouras, E., Santavas, N., Taitzoglou, A., Tarchanidis, K., Mitropoulos, A., and Gasteratos, A. (2019). Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations. *Sensors*, 19(16):3542.

[McKinney et al., 2011] McKinney, W. et al. (2011). pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9):1–9.

[Pires and Moreto, 2016] Pires, S. F. and Moreto, W. D. (2016). The illegal wildlife trade.

[Pouw and Dixon, 2020] Pouw, W. and Dixon, J. A. (2020). Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles. *Discourse Processes*, 57(4):301–319.

[Puri, 2021] Puri, A. (2021). A novel wildlife poaching detection solution using spatio-temporal data with dynamic time warping. *IEEE MIT Undergraduate Research Technology Conference, https://urtc.mit.edu/.*

[Puri and Bondi, 2021] Puri, A. and Bondi, E. (2021). Space, time, and counts: Improved human vs animal detection in thermal infrared drone videos for prevention of wildlife poaching. *ACM Knowledge Discovery and Data Mining (KDD) Conference FRAGILE EARTH Workshop, https://ai4good.org/fragile-earth-2021/.*

[Ramer, 1972] Ramer, U. (1972). An iterative procedure for the polygonal approximation of plane curves. *Comput. Graph. Image Process.*, 1:244–256.

[Ren and Hutchinson, 2008] Ren, L. and Hutchinson, J. R. (2008). The three-dimensional locomotor dynamics of african (loxodonta africana) and asian (elephas maximus) elephants reveal a smooth gait transition at moderate speed. *Journal of the Royal Society Interface*, 5(19):195–211.

[Sumpter, 2010] Sumpter, D. J. (2010). *Collective animal behavior.* Princeton University Press.

[Tavenard et al., 2020] Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., et al. (2020). Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, 21(118):1–6.

[Ten Holt et al., 2007] Ten Holt, G. A., Reinders, M. J., and Hendriks, E. A. (2007). Multidimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, volume 300, page 1.

[UN, 2021] UN (2021). United nations department of economic and social affairs sustainable development, https://sdgs.un.org/goals/goal15.

[US Congress Report, 2012] US Congress Report (2012). Ivory and insecurity: The global implication of poahing in africa, hearing before the committee on foreign relations united states senate, https://www.govinfo.gov/content/pkg/chrg-112shrg76689/html/chrg-112shrg76689.htm.

[Vuuren et al., 2019] Vuuren, M., Vuuren, R., Lutz, G., and Silverberg, L. (2019). On the effectiveness of uav for anti-poaching in the african arid savanna.

[Wildlife Crime, 2021] Wildlife Crime (2021). World wildlife fund: Wildlife crime technology project, https://www.worldwildlife.org/projects/wildlife-crime-technology-project.

[Wilson et al., 2015] Wilson, R. P., Griffiths, I. W., Mills, M. G., Carbone, C., Wilson, J. W., and Scantlebury, D. M. (2015). Mass enhances speed but diminishes turn capacity in terrestrial pursuit predators. *Elife*, 4:e06487.

[WWF, 2019] WWF (2019). World wildlife fund says african elephants will be extinct by 2040 if we don't act right away, Newsweek.

[WWF, 2021] WWF (2021). World wildlife fund: Elephant poaching facts, https://www.worldwildlife.org/species/elephant.